



QUATRA

21/11/2012
Delft

Software and Services for the Quality Management for Traffic Data (QUATRA)

Duration: 01/10/2011 – 30/09/2013 (24 months)

Budget: EUR 290.000

Coordinator: nast consulting ZT GmbH, Austria
Friedrich Nadler
e-mail: office@nast.at

Partner: Transver GmbH, Germany
Martin Rose
e-mail: rose@transver.de

Content of the presentation

1. main project objectives (reminder)
2. progress report on research method
3. results of the literature review
4. works carried out (model and software development, workshops)
5. next steps

Development of two tools for quality management

1. procedures and software tools will be developed to measure and estimate online the quality of incoming traffic data on **freeways** (for control centres)
2. furthermore a comparable service – offline – will be developed for **urban traffic** data (for cities and transport authorities)

The research method includes the following steps

- state-of-the art analysis
- development of the strategy for data evaluation
- development of algorithms and logic enquiries
- tests with historical data
- software development (applications, architecture)
- field tests
- analysis of test results



ongoing

not yet started

Results literature review

- objective: input of existing data quality evaluation for QUATRA
 - technical approaches (check of specific data features)
 - traffic engineering approaches (traffic flow fundamentals)
 - guidelines and standard procedures (specifications)
 - standardised systems and tools (developed/used at present)
-
- several interesting approaches identified
 - integration of tests and methods depends on available data (levels)
 - details included in development of indicators and criteria

Works carried out

- development of the strategy for data evaluation
- definition of indicators and criteria
- finalisation of test sections and data sources (freeways)
- workshops with road authorities
- development of statistical and logic procedures
- development of the software platform (until 06/2013)
- selection of use cases for the freeway tool

Strategy for data evaluation

- combination of indicators and statistical model
(Local/Global/Plausibility indicators)
- identification of “incidents”
- substitution of erroneous data based on historic traffic data

Selection of indicators and criteria

- two stage process
- first stage: project partners evaluated results of literature review separately
- second stage: interpretations and findings of both partners were shared and discussed
- definition of joined approach for definition of quality checks and model development

Example

Description, sources and suitability

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
1.	comparison of volumes and volume/occupancy ratio based on a 20-second intervals	Nihan Wang (1995)	volumes vs. detector utilisation - example: low traffic volumes should be represented in a low detector utilisation rate	1	0	1			
2.	vehicles change lanes physically at the location of the detection (splashover)	Coifman Lee (2011)	detectors on adjacent traffic lanes detect the same vehicle in case of lane change - could be filtered out (IF speed, time, vehicle is same)	2	1	1			
3.	same vehicles are detected twice at the same site (e.g. heavy vehicles with trailers)		testing of time gap at high speed detection, logical enquiry included in item M10	1	0	1			
4.	average vehicle lengths calculated and vehicle distributions estimated and compared with historical data	Turochy Smith (2000)							
5.	on-times were assessed (ratio of vehicle lengths over speed), vehicle distributions were logically compared with similar vehicle distributions and the average ontime for a time interval e respectively to identify faulty data.	Coifman Lee (2008) Chen May (1987)	percentage of vehicle categories on total traffic ok or wrong (differences during time/day/location)	1	1	1	1	0	0
6.	storage rates for time intervals	Nihan (1997)	can data be obtained about storage rates from single vehicle detections? Could be used as indicator for traffic queuing, logical enquiry included in item M19	1	1	1			
7.	minimum and maximum flow thresholds	Weijermars Berkum (2008)	included in items M12-M14						

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
8.	linear regression of the volumes of neighbouring sites	Chen et al. (2003)	for freeway section this criteria will be incorporated in the statistical model, for the urban road environment the density of traffic detection sites is highly relevant, high coverage is needed in order to get good results (partially included in item M20)	1	2	1	1	1	0
9.	measured flow values on neighbouring data collection sites								
10.	observation of unexpected high jumps of speed level and speed jumps			1	1	1	1	0	0
11.	If the speed is unexpected small, it is tested whether the speed is due to a traffic incident or measurement error occurred. The assumption was made that traffic incident in contrast to a measurement error is usually observed on all lanes.	Hoops (2002)	evaluation of average speeds in comparison to traffic volumes, indicator for traffic queuing	1	0	1	1	0	0
12.	Travel times from v and q		indicator for traffic queuing, can only be obtained in Austria through toll system data, for urban regions maybe Decell or INRIX data is available	1	2	1	1	0	0
13.	Under the assumption that the travel time of a considered and specified driver group stays stable, the travel time, calculated from the distance between the detectors and the measured speed, has to be similar to the travel time found by correlation analysis.		indicator for traffic queuing, can only be obtained in Austria through toll system data, for urban regions maybe Decell or INRIX data is available	1	2	1	1	2	
14.	conservation of flow: the test takes time series of different detection sites into account and calculates the quotient of the traffic volumes, estimates the distribution of the quotients and establishes the decision rules for data validation	nast consulting, TU Vienna (2008)	included in statistical model	1	2	1	1	1	0

Test sections on freeways

- Austria: sections on S 1, A 12
- Germany: sections on A 8, A 9, A 99
- Switzerland: no test sections

Test sections in urban areas

- Austria: Vienna
- Germany: Bremen

Workshops with authorities

- Toll system data could be used for evaluation purposes
- Check and substitution of speed data vital for traffic guidance
- User friendliness of the product very important
- Automated expert system could provide major benefit
- Labelling of erroneous data for statistics very important
(also in case of road works, accidents...)
- Online focus very important
- Interface for road works database would be beneficial

Local indicators

- Missing data: number or ratio of missing data sets
- Failure messages : number or ratio of data sets with failure message "255" generated by the detector itself

Global indicators

- Conservation of flow: ratio of number of vehicles at neighbouring measurement cross sections under consideration of inflow and outflow at ramps (different vehicle categories)

Plausibility indicators

- M1: $QKfz = 0 \Rightarrow (QLkw = 0 \text{ UND } QPkw = 0)$
- M2: $QKfz - QLkw = 0 \Rightarrow (QPkw = 0 \text{ UND } VPkw = 255)$
- M3: $QLkw = 0 \Rightarrow VLkw = 255$
- M4: $QPkw = 0 \Rightarrow VPkw = 255$
- M5: $QKfz \geq QLkw$
- M6: $QKfz - QLkw > 0 \Rightarrow 0 < VPkw$
- M7: $QKfz > 0 \Rightarrow 0 < VKfz$
- M8: $QLkw > 0 \Rightarrow 0 < VLkw$
- M9: $0 < t < T$
- M10: $QKfz = 0 \Rightarrow 0 < VGrenz,Kfz(t) = VGrenz,Kfz(t-T)$

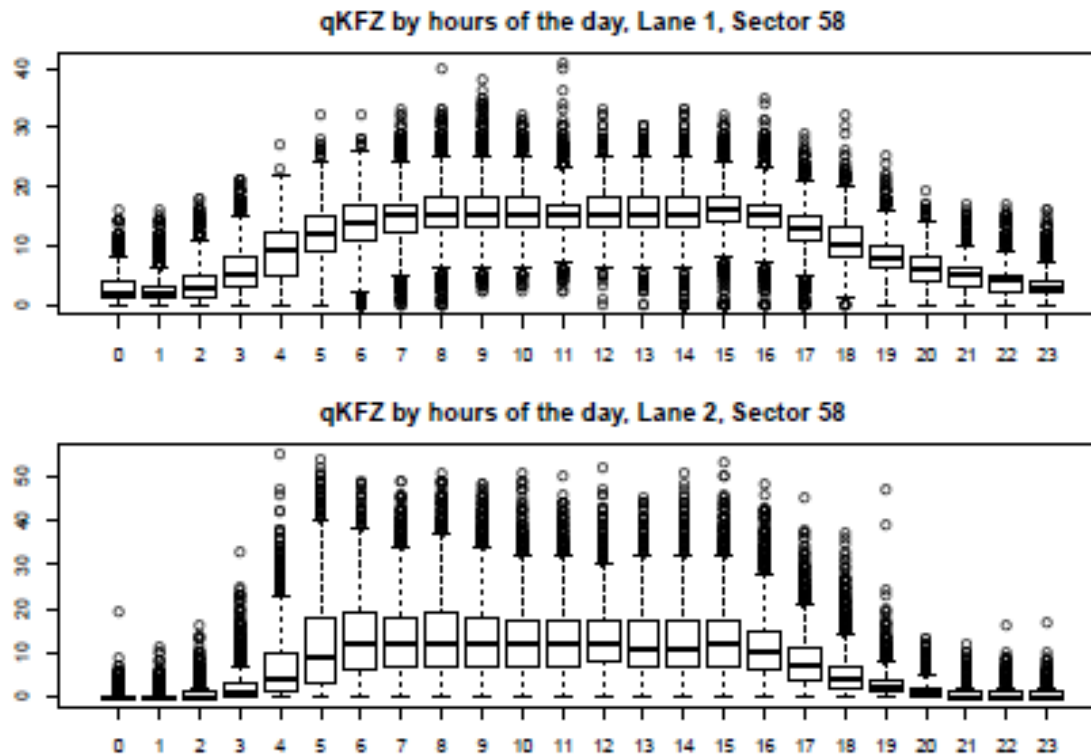
Plausibility indicators

- M11: $VKfz > VGrenz \Rightarrow B < Bgrenz$
- M12: $QKfz,min \leq QKfz \leq QKfz,max$
- M13: $QPkw,min \leq QPkw \leq QPkw,max$
- M14: $QLkw,min \leq QLkw \leq QLkw,max$
- M15: $VKfz,min \leq VKfz \leq VKfz,max$
- M16: $VLkw,min \leq VLkw \leq VLkw,max$
- M17: $VPkw,min \leq VPkw \leq VPkw,max$
- M18: $VGrenz,Kfz,min \leq VGrenz,Kfz \leq VGrenz,Kfz,max$
- M19: $Bmin \leq B \leq Bmax$
- M20: $VPkw,links > VPkw,rechts$
- M21: $VAusfahrt < VAusfahrt,grenz$
- M22: $QLkw,rechts > QLkw,links$

Statistical model

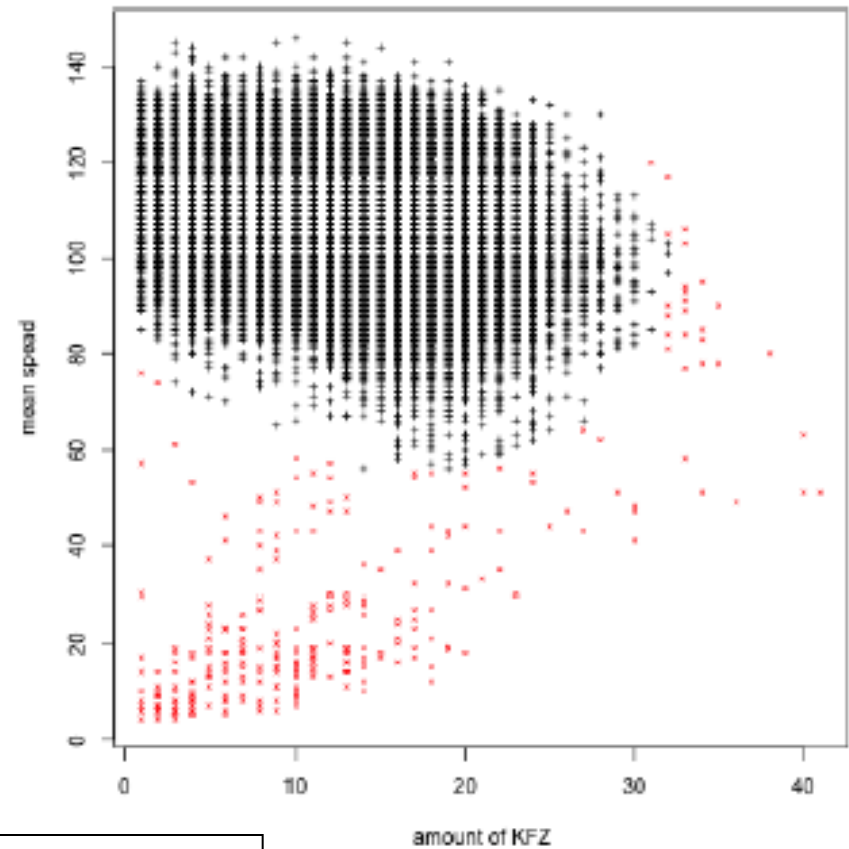
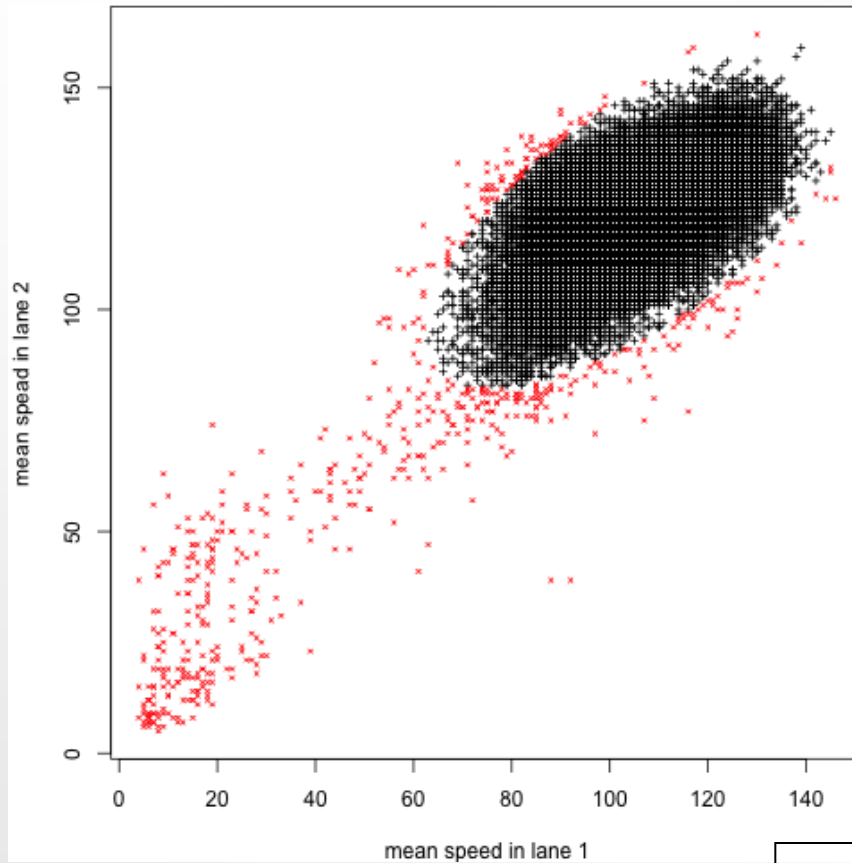
- In order to decide whether detector measurements are likely to be faulty a general regression model is developed
- Statistical information of historical data is generated and used
- Prediction intervals are generated
- Value prediction based on linear combination of independent variables of the data set
- Optional: pre-imputation of missing values for both the dependent variable as well as independent variables

Statistical model – data analysis



source: nast consulting

Statistical model – data analysis



source: nast consulting

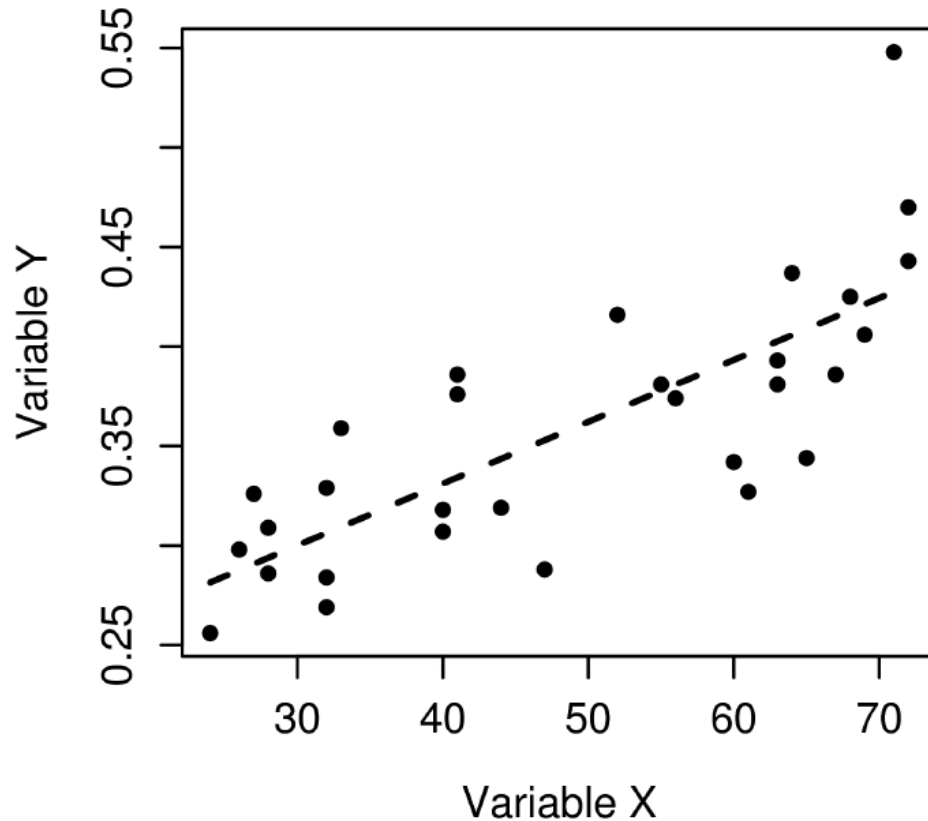
Statistical model – example for input data

- *log(tNetto)*: logarithm of variable net time gap
(mean net gap between vehicles measured in 1/10 seconds)
- *classVA*: indicator of traffic volume defined as ranges
- *Weekday and hours*: factor variable specifying weekday
(monday to sunday)
- *vPKW*: mean velocity of cars
- *lane*: indicator of lane
- *sector*: indicator of sector
- *Ist*: information provided by overhead displays
(hazard, queuing, pollution, speed restrictions)

Example for enquiry code definition

$$qKFZ \sim \log(tNetto) + classVA + weekday + vPKW + lane + sector + hour + Ist$$

Statistical model



Source:
Hutcheson, G. D. (2011)
Ordinary Least-Squares Regression
In L. Moutinho and G. D. Hutcheson
The SAGE Dictionary of Quantitative
Management Research
Pages 224-228.

Statistical model - identification of detector failure

- observed and collected values of number of vehicles for a given time interval at a specific lane on a given sector as well as the predicted number of vehicles for the exactly same time, lane and sector are available
- Identification of detector failure based on two assumptions
 - 1) the observed (measured) number of vehicles must lie within its corresponding prediction interval given by the results of the regression model
 - 2) the observed (measured) number of vehicles must lie within the range of the 2.5% and 97.5% percentile (normal range) of the distribution of number of vehicles for a given sector, lane, hour and hour of the day based on historical information

Statistical model - additional features

- Additional check of robust measures of location and deviance: median values, robust measurement of dispersion for specific sector, lane and hour of the day

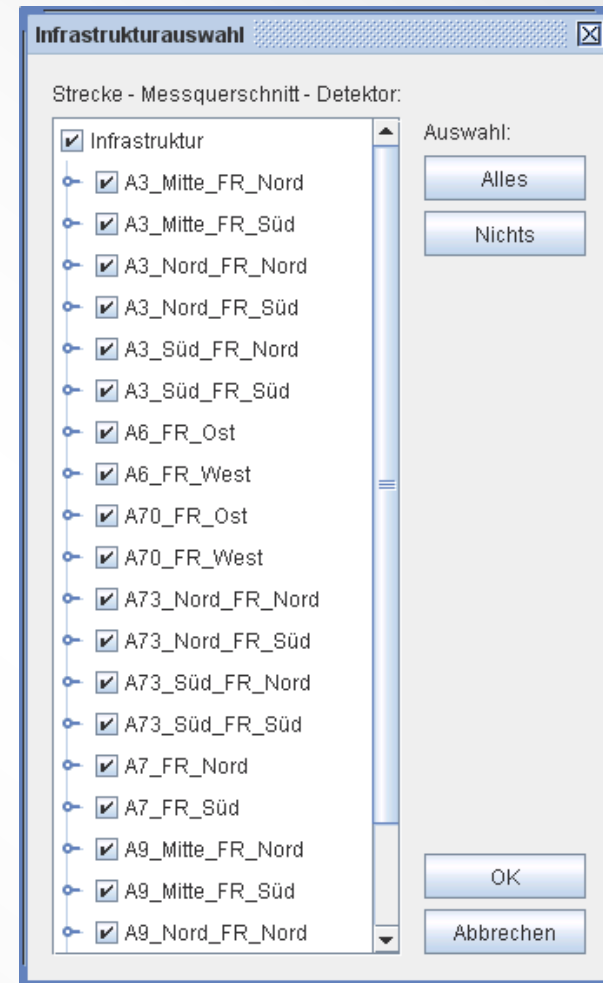
Software Development - basic functionalities

- Daily calculation of quality indicators for all relevant detection sites
- Clear visualisation of quality indicators
- Plot of tables and diagrams
- Labelling of erroneous data in accordance to defined thresholds
- Functions of the user interface
- Selection of locations (freeway, chainage, lane, detector)
- Time section
- Visualisation of the road sections (no. of lanes, ramps, detector location)

Basic functionalities

- Visualisation and export of indicators, tables und diagrams
- Dynamic window management (online tool)
- Selection of visualised indicators
- Display of thresholds
- Display of calculation parameters
- Diagram configuration
- Print, export, help, program settings

Location selection



Time selection

Zeitauswahl

Untersuchungszeitraum

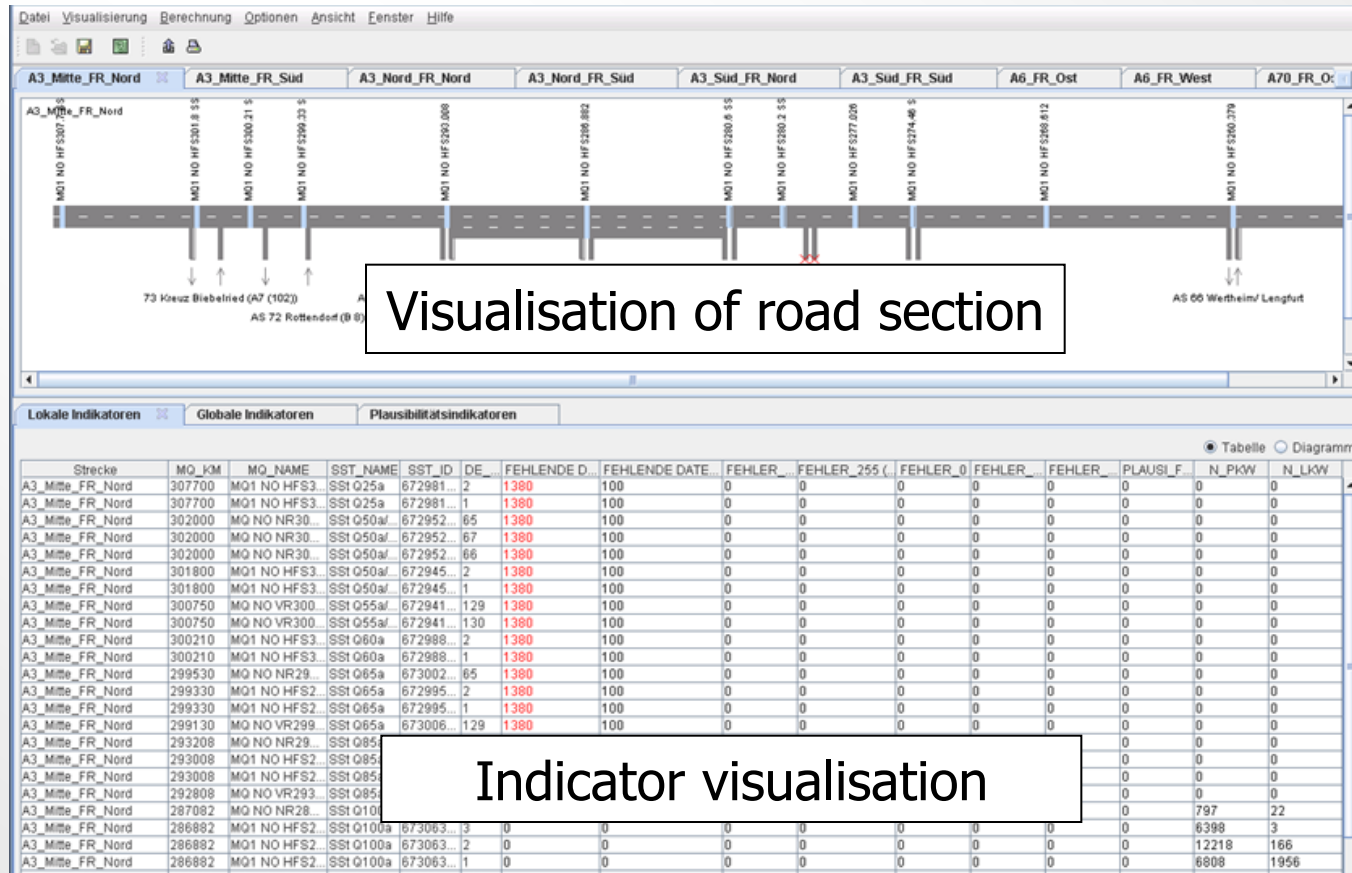
Datum: Uhrzeit (Stunde):
vom: 28.03.2010 00

Datum: Uhrzeit (Stunde):
August 2007 23

Abbrechen

comfortable calendar selection

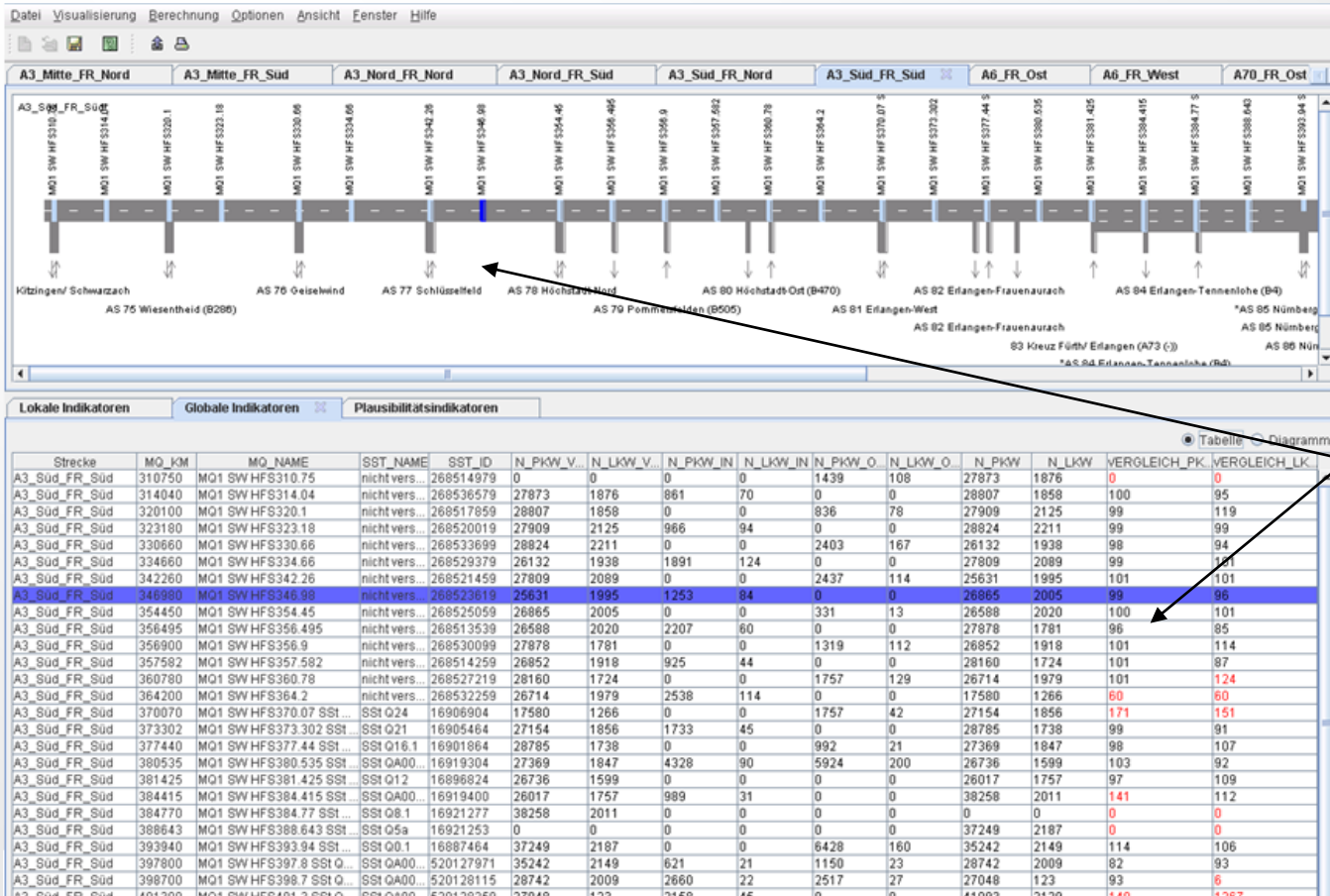
Indicators



Indicators

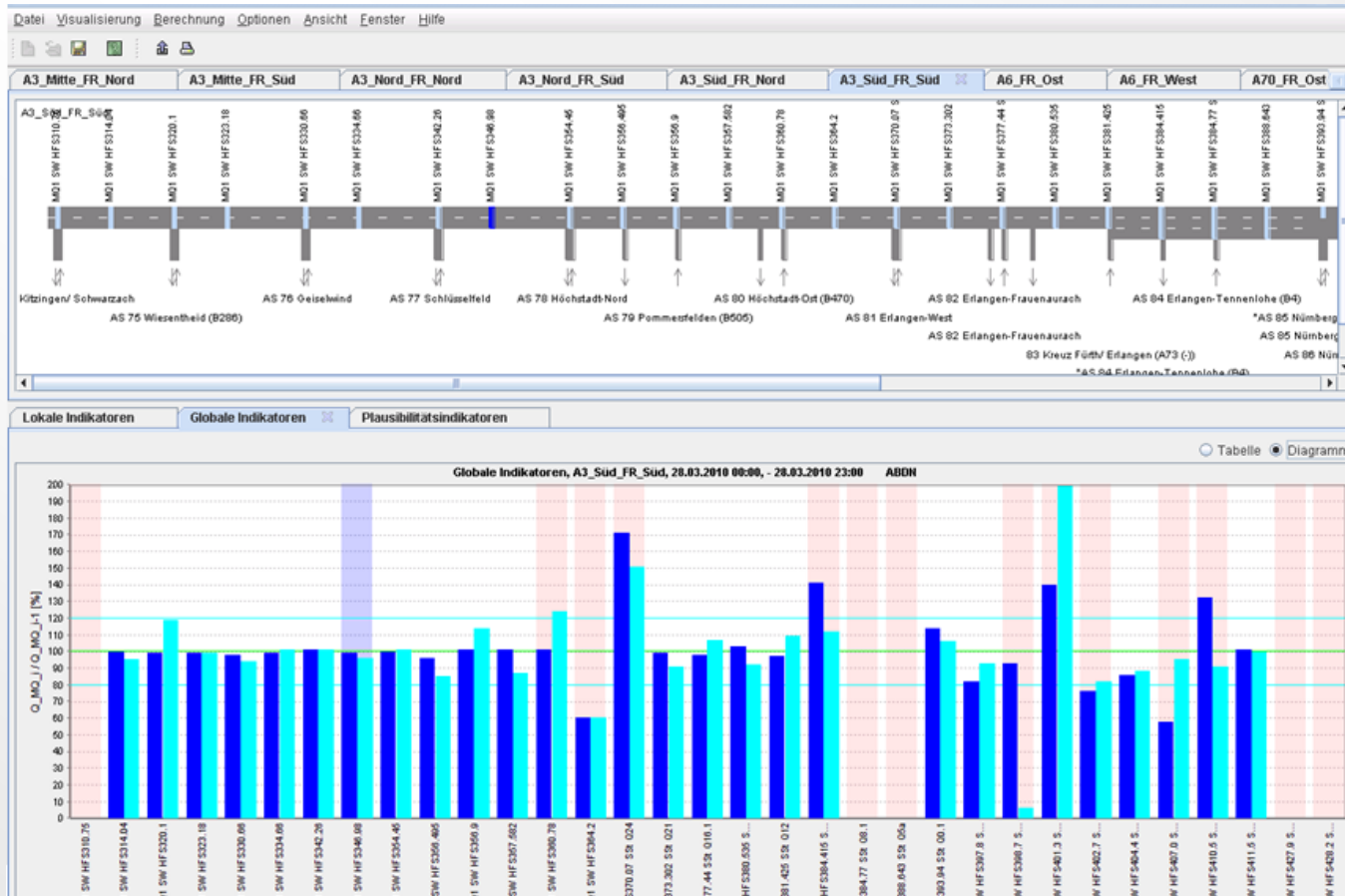


Indicators

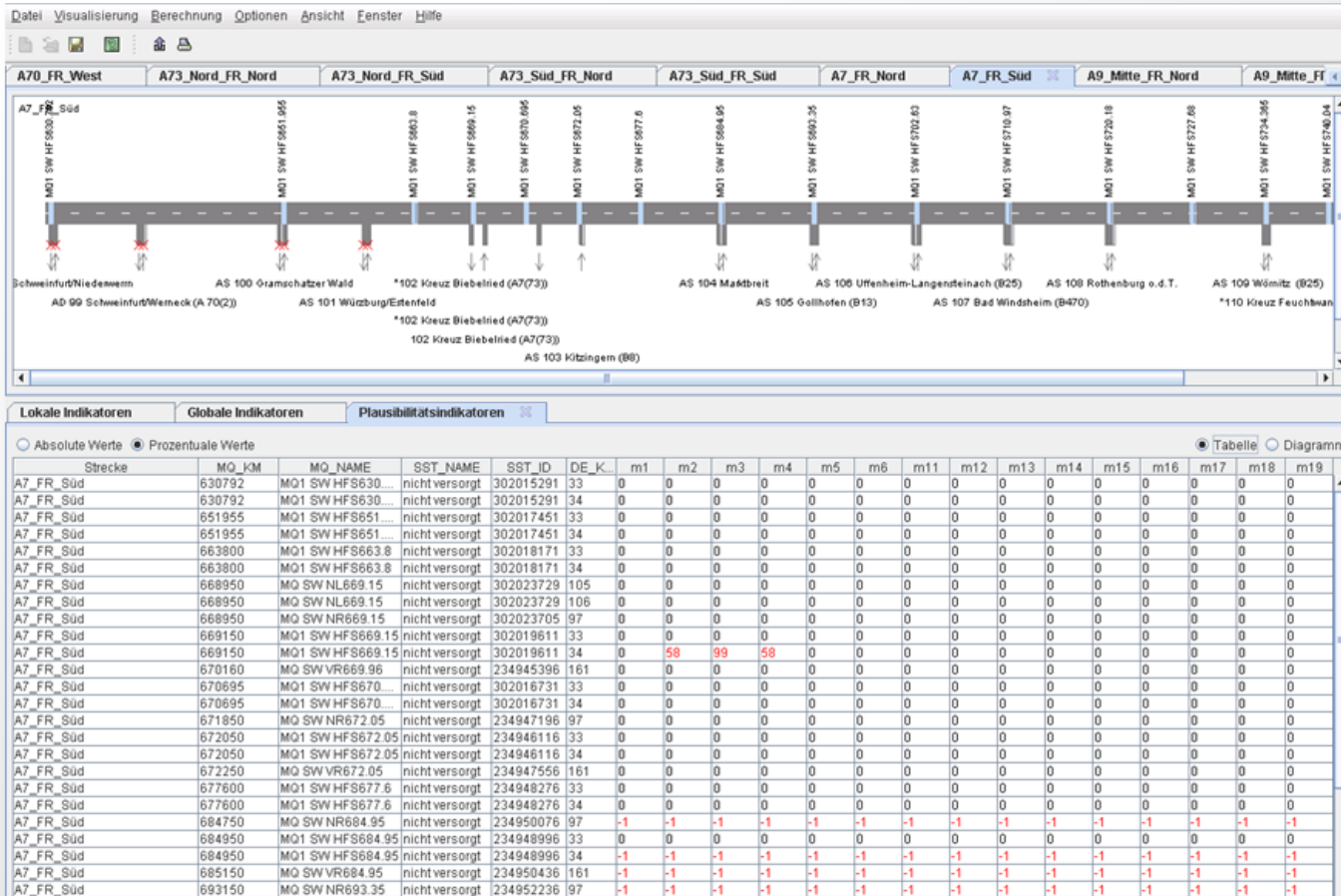


link of
information
on graph
and tables

Indicators



Indicators



Next steps

- Further testing of algorithms and indicators
- Further software development
- Intensive networking with Urban Road Authorities
- Comparison of error detection of QUATRA and existing systems

Expected results

QUATRA should be able to be used by various telematic operators to offer user friendly data check services with the required quality.

- quality management tools for the assessment of traffic data on freeways and in urban environments
- software which is able to import and use different kinds of data streams and formats
- the software will inform operators about the quality standard by e.g. issuing automated reports

Thank you for your attention

Daniel Elias

Tel: +43 1 523 47 33-28

elias@nast.at

www.nast.at

nast consulting

TRANSVER

Appendix

Statistical model

	possible suspicious	non-suspicious
Sept.	0.024	0.976
Oct.	0.030	0.970
Nov.	0.019	0.981
Dec.	0.021	0.979

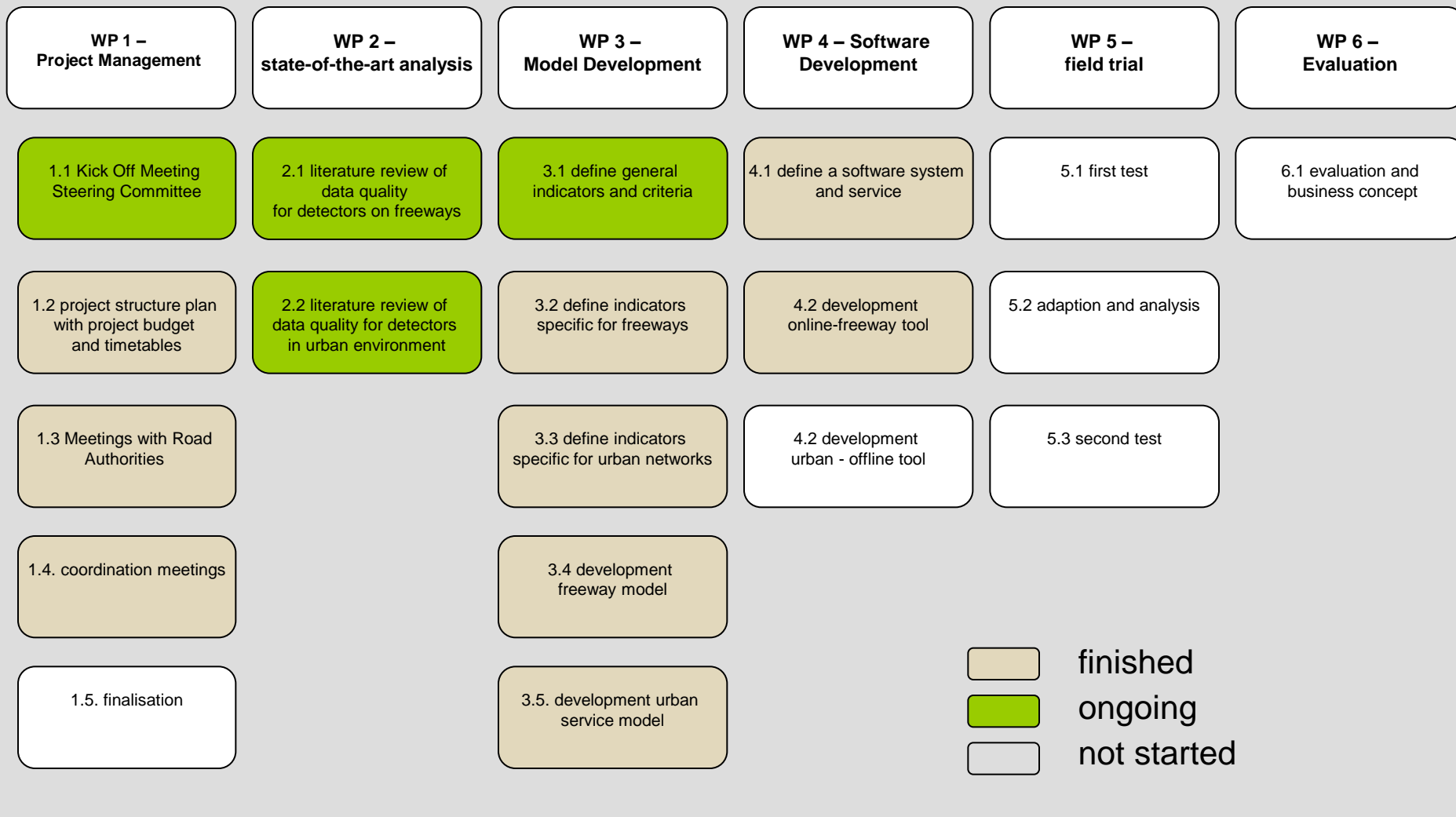
Table 9: Share of measurements classicied as suspicious and non-suspicious according to strategy based on prediction intervals and historical quantiles

	possible suspicious	non-suspicious
Sep.	0.011	0.989
Oct.	0.011	0.989
Nov.	0.011	0.989
Dec.	0.011	0.989

Table 10: Share of measurements classicied as suspicious and non-suspicious according to strategy based on standardized residuals and historical quantiles

	possible suspicious	non-suspicious
Sept.	0.06	0.94
Oct.	0.05	0.95
Nov.	0.05	0.95
Dec.	0.06	0.94

Table 11: Detection of possible wrong measurements using robust techniques



Item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)			global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)		toll system data, Bluetooth, ANPR, DECELL, INRIX	
15.	value occupancy and flow minimum and maximum thresholds	Freudenberger (2001)	included in items M12-14 and M19	1	1	1				
16.	The speed was calculated from q and o due to the correlation of the fundamental diagram. If the speed exceeded the speed limit at the specific detector, the dataset was marked as implausible		logical combination of parameters of fundamental diagram (can only be evaluated if individual vehicle speeds are available for analysis)	1	1	1	1	0	0	
17.	The data was checked regarding recurrence		check, if single vehicles can be detected (data issue)	1	1	1	1	1	1	
18.	regression curve second order was placed in a fundamental diagram. Measurement data with a significant high distance to the regression curve were marked as suspect		included in statistical model	1	1	1	1	0	0	
19.	The coefficient of determination described how good the scatter plot was represented by the regression curve. If the coefficient of determination was small, there was only a small correlation between o and q and the detector was treated as possibly defect.			1	1	1	1	1	0	
20.	conservation of flow: comparison of number of vehicles at neighbouring traffic sites - under consideration of on- and off ramps	nast consulting TRANSVER	included in statistical model	1	2	1	1	1	1	
21.	conservation of flow: comparison of number of heavy vehicles at neighbouring traffic sites - under consideration of on- and off ramps		included in statistical model	1	0	1	1	1	0	
22.	M1: $QKfz = 0 \Rightarrow (QLkw = 0 \text{ UND } QPkw = 0)$		total traffic must be the sum of individual vehicle categories. If no vehicles total were counted during time period no car or HV can be registered	1	1	1	1	0	0	

Item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)			global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)		toll system data, Bluetooth, ANPR, DECELL, INRIX	
23.	M2: $QKfz - QLkw = 0 \Rightarrow (QPkw = 0 \text{ UND } VPkw = 255)$	nast consulting TRANSVER	example: if total traffic = total vehicles category 1 there can't be any amount of vehicles for category 2... - furthermore speed of category 2... must be 0	1	1	1	1	0	0	
24.	M3: $QLkw = 0 \Rightarrow VLkw = 255$		if no HV has been counted average speed must be 0 or code 255	1	1	1	1	0	0	
25.	M4: $QPkw = 0 \Rightarrow VPkw = 255$		if no car has been counted average speed must be 0 or code 255	1	1	1	1	0	0	
26.	M5: $QKfz \geq QLkw$		total traffic must be sum of all vehicles categories	1	1	1	1	0	0	
27.	M6: $QKfz - QLkw > 0 \Rightarrow 0 < VPkw$		if cars have been counted average car speed must increase	1	1	1	1	0	0	
28.	M7: $QKfz > 0 \Rightarrow 0 < VKfz$		if vehicles have been counted average speed must be present	1	1	1	1	0	0	
29.	M8: $QLkw > 0 \Rightarrow 0 < VLkw$		if vehicles have been counted average speed must be present	1	1	1	1	0	0	
30.	M9: $0 < t < T$		covers item 1 - detector utilization time must be higher 0 and shorter than time interval	1	2	1	1	0	0	
31.	M10: $QKfz = 0 \Rightarrow 0 < Vg.Kfz(t) = Vg.Kfz(t-T)$		if no vehicle has been counted during time interval, averaged vehicle time must be higher 0 and same as previous time period	1	2	1	1	0	0	
32.	M11: $VKfz > VGrenz \Rightarrow B < BGrenz$		if average vehicle speed is high during a time interval the detector occupancy rate has to be below a certain threshold (fundamental diagram)	1	2	1	1	0	0	
33.	M12: $QKfz_{min} \leq QKfz \leq QKfz_{max}$		traffic volumes during a certain time have to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0	

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
34.	M13: QPk _{w,min} ≤ QPk _w ≤ QPk _{w,max}		car volumes during a certain time have to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0
35.	M14: QLk _{w,min} ≤ QLk _w ≤ QLk _{w,max}		HV volumes during a certain time have to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0
36.	M15: VKf _{z,min} ≤ VKf _z ≤ VKf _{z,max}		average vehicle speed during a certain time has to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q,v,k,s, occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loops/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
37.	M16: VLk _{w,min} ≤ VLk _w ≤ VLk _{w,max}	nast consulting TRANSVER	average HV speed during a certain time has to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0
38.	M17: VPk _{w,min} ≤ VPk _w ≤ VPk _{w,max}		average car speed during a certain time has to be within a certain range - otherwise there is a disturbance - refer to item 7	1	1	1	1	0	0
39.	M18: Vg,Kfz,min ≤ Vg,Kfz ≤ Vg,Kfz,max		smoothed vehicle speed from during a certain time has to be within a certain range - otherwise there is a disturbance	1	2	1	1	0	0
40.	M19: B _{min} ≤ B ≤ B _{max}		average car speed during a certain time has to be within a certain range - otherwise there is a disturbance - refer to item 8	1	2	1	1	0	0
41.	M20: VPk _{w,links} > VPk _{w,rechts}		Germany: average car speed in right freeway lane should be below average car speed in left freeway lane (due to driver behaviour this assumption can not be used for Austrian motorways and urban areas)	1	2	1	1	0	0
42.	M21: VAusfahrt < VAusfahrt.grenz		average vehicle speed at on/off ramps during a certain time has to be within a certain range - otherwise there is a disturbance	1	0	0	1	0	0
43.	M22: QLk _{w,rechts} > QLk _{w,links}	DAVINCI	HV volume in left freeway lane should be below HV volume in left freeway lane (problem during overtaking manoeuvres)	2	2	1	1	0	0
44.	Indication of sources of errors based on previous events of the same type that were saved in the database		should be included in a follow up project	0	0	0	-	-	-
45.	Peak-Hour-lane occupancy: erroneous data outside of operating hours		outside of operating hours the number of detected vehicles should be below a certain threshold	1	0	1	1	0	0
46.	Acceptance monitor of speeds (checking whether the traffic control strategy, which was shown on the road, was accepted by the drivers)	BUSCH	not part of core focus of quatra, could be used for evaluation of compliance to dynamic signage	0	0	0	-	-	-

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q.v.k.s. occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loop/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
47.	For the current project the possible fault scenarios and sources of error of the individual ways of data collecting described in the reference note are of interest.	FGSV AK 3.5.20	wrong number of vehicles or no vehicle recorded, wrong classification of vehicles, wrong speed, repetition of interval data, included in logical enquiries M	1	1	1			
48.	repetition of interval data with same time stamp	-	indicator for data error	1	1	1	1	1	0
49.	comparison of volumes and average speeds	DVS Netz pilot nast consulting	according to results of research project high chance of queuing in case of high volumes and low average speeds	1	1	1	1	0	0
50.	comparison of volumes and average speeds	DVS Netz pilot nast consulting	in case of high traffic density average speed can't be high due to fundamental diagram	1	1	1	1	0	0
51.	Highway Agencies' requirements for assessment of data quality (target parameter, degree of accuracy)	MCH1529	potential benefit for software platform	1	0	1	1	1	1
52.	fault monitoring/reporting		potential benefit for software platform	1	1	1	-	-	-
53.	Plausibility check of short-term data, differentiation between inflowing and outflowing traffic	FGSV AK 3.5.20	covered in item 10	2	1	0	-	-	-
54.	data quality requirements regarding completeness, availability, veracity, precision, timeliness	QUANTIS	included in some of the M rules				-	-	-
55.	missing data (quantities)	nast consulting TRANSVER	potential benefit for software platform output	1	1	1	1	1	1
56.	missing data (percentage)		potential benefit for software platform output	1	1	1	1	1	1
57.	missing data code 255 (quantities)		potential benefit for software platform output	1	1	1	1	0	0
58.	missing data code 255 (percentage)		potential benefit for software platform output	1	1	1	1	0	0
59.	missing data code 0 (quantities)		potential benefit for software platform output	1	1	1	1	1	1
60.	missing data code 0 (percentage)		potential benefit for software platform output	1	1	1	1	1	1
61.	flagged data sets (resulting out of quality checks - quantities)		potential benefit for software platform output	1	1	1	1	1	1
62.	flagged data sets (resulting out of quality checks - percentage)		potential benefit for software platform output	1	1	1	1	1	1

item	criteria and indicators	source	description	0 - not suitable, 1 - suitable, 2 - partially suitable			stationary and local data (q.v.k.s. occ...)		global data
				freeway	urban	temp hard shoulder	TLS data sources (loop/radar/TEU)	toll system data, bluetooth, ANPR, DECELL, INRIX	
63.	monitoring levels	TRAFFIC IQ	potential benefit for software platform	1	1	1			
64.	reference systems and data fusion		potential benefit for software platform	1	1	0	1	1	1
65.	examined under a full range of traffic (Free Flowing, Heavy traffic with flow breakdowns, Stop-start conditions, Varied classifications of vehicles) and environmental conditions likely to be experienced on the network	MCH1529	included in some of the M rules	1	1	1	1	1	1
66.	comparison of data detection values and interpolated data (Quatra output)	workshop ASFINAG 01/2012	potential benefit for software platform output	1	1	1	1	1	1